


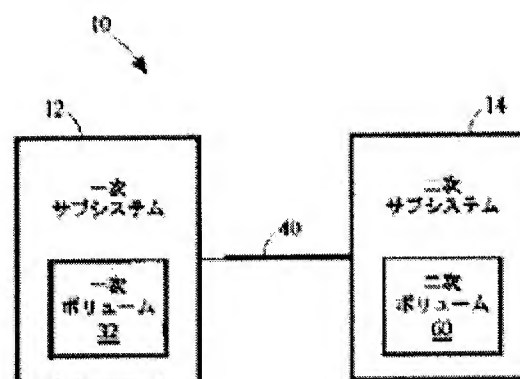
DEVICE AND METHOD FOR PERFORMING SNAP SHOT COPY OF VOLUME**Publication number:** JP11259348**Publication date:** 1999-09-24**Inventor:** CHRISTOPHER JAMES WEST; DAVID ALAN SEARLES**Applicant:** IBM**Classification:****- international:** G06F12/00; G06F11/14; G06F11/20; G06F12/00; G06F11/14; G06F11/20; (IPC1-7): G06F12/00**- European:** G06F11/14A4B1; G06F11/14A4B1M8**Application number:** JP19990009139 19990118**Priority number(s):** US19980013712 19980126**Also published as:** US6131148 (A1)

Report a data error here

Abstract of JP11259348

PROBLEM TO BE SOLVED: To provide a method and a device for performing the snap shot copy of a remote secondary volume to the other volume on a remote subsystem by setting up a peer-to-peer remote copy (PPRC) session.

SOLUTION: This device is provided with a primary storage subsystem 12 having a primary data storage device equipped with at least one primary volume 32. A primary processor repeats a request for the execution of snap shot copy at one part of the primary volume 32 at least to a secondary storage subsystem 14. The secondary storage subsystem 14 is provided with a secondary data storage device having a secondary volume 60 and while using the remote copy session, the secondary volume 60 is kept duplexed with the primary volume 32. A secondary processor performs the snap shot copy of the correspondent section of the secondary volume 60 in response to the repeated request so as to perform operation equal with the snap shot copy at a part of the primary volume 32 at least.



Data supplied from the esp@cenet database - Worldwide

(19) 日本国特許庁 (J P)

(12) 公開特許公報 (A)

(11) 特許出願公開番号

特開平11-259348

(43) 公開日 平成11年(1999) 9月24日

(51) Int.Cl.⁶

G 0 6 F 12/00

識別記号

5 3 1

F I

G 0 6 F 12/00

5 3 1 M

5 3 1 D

審査請求 未請求 請求項の数13 O L (全 10 頁)

(21) 出願番号 特願平11-9139

(22) 出願日 平成11年(1999) 1月18日

(31) 優先権主張番号 0 9 / 0 1 3 7 1 2

(32) 優先日 1998年 1月26日

(33) 優先権主張国 米国 (US)

(71) 出願人 390009531

インターナショナル・ビジネス・マシーンズ・コーポレーション

INTERNATIONAL BUSINESS MACHINES CORPORATION

アメリカ合衆国10504、ニューヨーク州
アーモンク (番地なし)

(72) 発明者 クリストファー・ジェームズ・ウェスト

アメリカ合衆国80303 コロラド州ボルダー
エルムズハースト・プレース 3860

(74) 代理人 弁理士 坂口 博 (外1名)

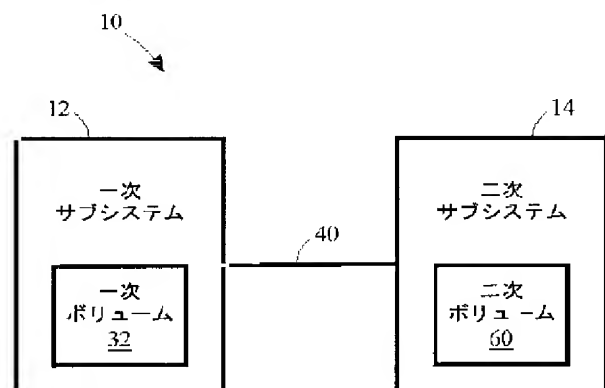
最終頁に続く

(54) 【発明の名称】 ボリュームのスナップショット・コピーを行う装置及び方法

(57) 【要約】

【課題】 本発明は、ピア・ツー・ピア遠隔コピー (PPRC) セッションをセットアップし、遠隔二次ボリュームを遠隔サブシステム上の他のボリュームにスナップショット・コピーする方法及び装置を提供する。

【解決手段】 この装置は、少なくとも1つの一次ボリュームを備える一次データ記憶装置を有する一次記憶サブシステムを含む。一次処理装置が、一次ボリュームの少なくとも一部のスナップショット・コピーの実行を求める要求を二次記憶サブシステムに中継する。二次記憶サブシステムは二次ボリュームを有する二次データ記憶装置を含み、二次ボリュームは遠隔コピー・セッションを使用して一次ボリュームとの二重化状態に維持されている。二次処理装置が、中継された要求に応答して、二次ボリュームの対応する部分のスナップショット・コピーを行うことによって一次ボリュームの少なくとも一部のスナップショット・コピーと同等の操作を行うことができる。



【特許請求の範囲】

【請求項1】ボリュームのスナップショット・コピーを行う装置であって、一次記憶サブシステムと二次記憶サブシステムとを含み、前記一次記憶サブシステムは、

(a) 少なくとも1つの一次ボリュームを有する一次データ記憶装置と、

(b) 前記一次ボリュームの少なくとも一部のスナップショット・コピーを行う要求を二次記憶サブシステムに中継する一次処理手段とを含み、

前記二次記憶サブシステムは、

(c) 前記一次ボリュームの複製コピーである二次ボリュームを有する二次データ記憶装置と、

(d) 前記要求にตอบสนองして、前記二次ボリュームに対応する部分のスナップショット・コピーを行うことによって前記一次ボリュームの前記少なくとも一部のスナップショット・コピーと同等のものを実行する二次処理手段とを含む装置。

【請求項2】前記二次ボリュームが前記一次ボリュームの複製コピーになるように、前記一次ボリュームの前記二次ボリュームへのバックアップ・コピー操作を実行する手段をさらに含む、請求項1に記載の装置。

【請求項3】前記バックアップ・コピー操作が進行中でない場合にのみ、前記一次処理手段が前記要求を中継する、請求項2に記載の装置。

【請求項4】前記スナップショット・コピーが第3の遠隔ボリュームに記憶される、請求項1に記載の装置。

【請求項5】前記二次ボリュームの前記少なくとも一部の前記スナップショット・コピーが前記二次記憶サブシステムの仮想メモリに記憶される、請求項1に記載の装置。

【請求項6】前記一次記憶サブシステムが、前記一次ボリューム上のデータの物理場所を仮想アドレスに関連づける一次ディレクトリ情報を含む、請求項1に記載の装置。

【請求項7】前記一次ディレクトリ情報がログ構造配列形式である、請求項6に記載の装置。

【請求項8】前記二次記憶サブシステムが、前記二次ボリューム上のデータの物理場所を仮想アドレスに関連づける二次ディレクトリ情報を含む、請求項1に記載の装置。

【請求項9】前記二次ディレクトリ情報がログ構造配列形式に構成されている、請求項8に記載の装置。

【請求項10】前記スナップショット・コピーが完了すると、前記二次記憶サブシステムが前記スナップショット・コピー操作の完了を示す信号を前記一次記憶サブシステムに送る、請求項1に記載の装置。

【請求項11】ボリュームのスナップショット・コピーを行う方法であって、一次記憶サブシステムの一次ボリュームの少なくとも一部のスナップショット・コピーを行う要求を前記一次記

憶サブシステムで受け取るステップと、

前記要求を前記一次記憶サブシステムから、前記一次ボリュームの複製である二次ボリュームを有する二次記憶サブシステムに中継するステップと、

前記二次ボリュームの対応する部分についてスナップショット・コピーを行うことによって前記一次ボリュームの前記少なくとも一部のスナップショット・コピーと同等のものを生成するステップとを含む方法。

【請求項12】前記二次ボリュームが前記一次ボリュームの複製になるように、前記一次記憶サブシステムが、前記二次ボリュームへの前記一次ボリュームのバックアップ・コピー操作を行う手段を含む、請求項11に記載の方法。

【請求項13】前記バックアップ操作が進行中でない場合に前記生成ステップが行われる、請求項11に記載の方法。

【発明の詳細な説明】**【0001】**

【発明の属する技術分野】本発明は、一次記憶サブシステムと遠隔二次記憶サブシステムとの間で遠隔二重コピー技法を使用するデータ記憶システムのボリュームのスナップショット・コピーに関し、より詳細には、遠隔二次データ記憶システムにおいて一次記憶システム内のボリュームのスナップショット・コピーと同等の操作を実行することに関する。

【0002】

【従来の技術】周辺データ記憶装置などのデータ処理システムに記憶されているリアルタイム・バックアップ保護のために、データのローカルまたは遠隔あるいはその両方のコピー機能を備えることが重要である。遠隔データ・センターへのバックアップまたはコピーによって、ローカル・バックアップ・コピーで保護されていない物理災害から保護される。この自動データ・コピーを一般に、遠隔二重化または遠隔データ・コピーと呼ぶ。従来の技術は、復元とアーカイブのためにデータをバックアップする方法及び手段を十分に備えている。このようなシステムの一例は、デュヤノビッチ (Duyanovich) 等の米国特許第555371号に記載されている。

【0003】データの保存と保護のためにバックアップ・コピー技法が頻繁に使用されるが、複数の遠隔または二次バックアップ記憶サイトの使用はコストがかなり高くつき、大量のデータを記憶するデータ記憶システムの場合は特にそうである。したがって、コスト効果の高いデータ保存技法を提供する必要がある。

【0004】よりコスト効果の高いデータ保存技法は、スナップショット・コピーを行うことである。スナップショット・コピーとは、(複数のボリュームを含むことがある) データ記憶装置内のデータの物理場所またはディレクトリの(すなわち物理アドレス、サイズなど)のイメージである。スナップショット・コピーは、万一デ

ータ記憶装置に障害が発生した場合、保護されたデータ記憶装置のイメージを復元するため、または前の指定期間からのデータの履歴スナップショットをそのデータの現行バージョンと比較するために使用することができる。スナップショット・コピーの利点は、どのデータ・ファイル、ボリュームなどのスナップショット・コピーにも数秒でアクセスすることができることと、スナップショット・コピーによってデータ・ファイルまたはボリュームを、スナップショット・コピーがとられた瞬間の元のバージョンに復元することができることである。

【0005】スナップショット・コピーの1つの問題は、スナップショット・コピー中のボリュームへのアクセスが、スナップショット操作が完了するまで延期されることである。これによって、スナップショット操作の実行中に同時にそのボリュームへのアクセスや変更が行われないように防止する。他の問題は、スナップショット・コピーでは、スナップショット・コピーのために記憶空間やアドレス指定を割り振る必要があることである。これは、記憶容量が限られていたりアドレスが限られているデータ・システムでは特に問題である。したがって、ユーザはデータ記憶システムの効率、アクセス可能性、及び記憶容量に悪影響を与えずに行うことができるスナップショット操作の数が制限される。

【0006】スナップショット・コピーのために一次記憶サブシステムで記憶空間を割り振る必要なしに、一次記憶サブシステム上のボリュームをスナップショット・コピーすると同時に、スナップショット・コピー中のボリュームにアクセスできるようにする方法及び装置も必要である。

【0007】

【発明が解決しようとする課題】したがって、本発明の目的は、一次データ記憶装置で記憶空間を割り振る必要なしに、一次データ記憶サブシステムのボリュームのスナップショット・コピー操作と同等の操作を行う方法及び装置を提供することである。

【0008】本発明の他の目的は、一次データ記憶サブシステムのボリュームのスナップショット・コピーと同等の操作を実行すると同時に、スナップショット・コピー操作の実行中にホスト装置がそのボリュームにアクセスすることができるようにする方法及び装置を提供することである。

【0009】本発明の他の目的は、二次データ記憶サブシステムを使用して一次データ記憶サブシステムのボリュームのスナップショット・コピーと同等の操作を行う方法及び装置を提供することである。

【0010】本発明の他の目的は、ボリュームの遠隔二重コピーをすでに行っている既存のデータ記憶システムに組み込むことができる一次データ記憶装置のボリュー

ムのスナップショット・コピーを行う方法及び装置を提供することである。

【0011】

【課題を解決するための手段】本発明は、ピア・ツー・ピア遠隔コピー（PPRC）セッションをセットアップし、遠隔二次ボリュームを遠隔サブシステム上の他のボリュームにスナップショット・コピーする方法及び装置を提供する。この装置は、少なくとも1つの一次ボリュームを備える一次データ記憶装置を有する一時記憶サブシステムを含む。一次処理装置が、一次ボリュームの少なくとも一部のスナップショット・コピーを行う要求を二次記憶サブシステムに中継する。二次記憶サブシステムは、遠隔コピー・セッションの使用により一次ボリュームとの二重化状態に維持されている二次ボリュームを有する二次データ記憶装置を含む。二次処理装置が、中継された要求にตอบสนองして、二次ボリュームの対応する部分のスナップショット・コピーを作成することによって二次ボリュームの少なくとも一部のスナップショット・コピーと同等の操作を行うことができる。

【0012】二次ボリュームのスナップショット・コピーが特定の時点で一次ボリュームのスナップショット・コピーと同等であるように保証するため、一次ボリュームと二次ボリュームを二重化状態に維持することが重要である。これは、一次ボリュームが変更されると二次ボリュームと二次ディレクトリ情報を更新することによって行う。したがって、二次ボリュームの更新中に二次ボリュームのスナップショット・コピー操作が行われないことが好ましい。すなわち、一次サブシステムは、一次ボリュームと二次ボリュームが確実に二重化状態になっていることを確認してからスナップショット・コピー操作を行うことができるようにし、それによって誤ったデータがコピーされるのを防ぐ。

【0013】本発明の機構は、一次ボリュームのスナップショット・コピー操作の実行に伴う一次サブシステムにおける遅延と記憶域の問題を回避する。すなわち、二次サブシステムでスナップショット・コピー操作を行って同じ結果、すなわち一次ボリュームのスナップショット・コピーと同等のものを得ることができる。本発明のスナップショット・コピー機能は、ボリュームをバックアップするためにすでにピア・ツー・ピア遠隔コピー（PPRC）技法を使用している記憶システムに組み込むこともできる。

【0014】

【発明の実施の形態】図面の詳細な説明に入る前に、本明細書で使用する特定の用語を定義しておいた方がわかりやすい。以下、「バックアップ・コピー」という用語は、遠隔二次データ記憶サブシステム上で一次データ記憶サブシステムの一次ボリュームの複製コピーを作成する操作を指す。バックアップ・コピーに付随する他の用語として、二重化、二重コピー、ミラーリングなどがあ

り、これらはすべて当技術分野で一般に知られている。ピア・ツー・ピア遠隔コピー（PPRC）セッションまたは遠隔バックアップ・コピー・セッションとは、一次ボリューム内の更新データを二次ボリュームにバックアップ・コピーするシーケンスの実施を指す。

【0015】「遠隔」という用語は、一次ボリューム及び二次ボリュームの場所を限定するものと解釈してはならない。一次サブシステムと二次サブシステムは、異なる場所にあっても同じデータ・センター内にあってもよい。また、一次及び二次データ記憶システムは互いに独立していてもよく、同一または異なるホスト・プロセッサに接続されていてもよい。

【0016】ピア・ツー・ピア遠隔コピー（PPRC）とは、一次サブシステムの一次ボリュームと、二次サブシステムなど遠隔記憶場所にある一次ボリュームの複製コピー、すなわち二次ボリュームとを指す。PPRCは、遠隔バックアップ・コピーセッションの使用により二重化状態に維持される。説明をわかりやすくするために、以下、PPRCは一次サブシステムの一次ボリュームと遠隔二次サブシステムの二次ボリュームとを指すものとする。

【0017】ボリュームのスナップショット・コピーとは、ボリュームの論理ディレクトリのコピーを指す。論理ディレクトリはデータをボリュームの物理場所、すなわちトラックに関連づける。スナップショット・コピーは、スナップショット・コピーが行われた特定の時点のボリュームのイメージ、詳細にはボリューム内のデータの構成を提供する。ボリュームのスナップショット・コピーによって、ボリュームを再構成したり、スナップショット・コピーが行われた時点におけるボリュームの状態と比較したりすることができる。

【0018】本発明の詳細な説明に移ると、図1には、通信インタフェース40を介して接続された一次サブシステム12と少なくとも1つの遠隔二次サブシステム14を含むデータ・システム10の全体構造図が図示されている。一次サブシステム12は、一次ボリューム32内にローカルに記憶されているデータの受信、生成、または変更を行う。障害時の復元のために、一次ボリューム32内のデータは、遠隔バックアップ・コピー・セッションの使用によって二次サブシステム14の複製二次ボリューム60にバックアップされる。したがって、一次サブシステム12と二次サブシステム14は協調して、バックアップのために自動的にデータを転送し、それによって一次ボリューム32と二次ボリューム60が二重化状態になるようにする。

【0019】一次ボリューム32のスナップショット・コピー操作の実行に伴う一次サブシステム12での遅延と記憶域の問題を回避するために、スナップショット・コピー操作を二次サブシステム14で行って同じ結果、すなわち一次ボリューム32のスナップショット・コピ

ーと同等のものを得ることができることがわかった。すなわち、二次サブシステム14内には一次ボリューム32の複製コピー、すなわち二次ボリューム60が記憶されている。したがって、二次サブシステム14は、二次ボリューム60に対してスナップショット・コピー操作を行うことによって一次ボリューム32のスナップショット・コピーと同等のコピーが得られる。

【0020】動作中、一次サブシステム12はホスト装置またはユーザ（図示せず）から、一次ボリューム32に対してスナップショット・コピー操作を実行するように求めるスナップショット要求を受け取る。一次サブシステム12は、このスナップショット・コピー要求を二次サブシステム14に中継または転送する。その後、二次サブシステム14は二次ボリュームのスナップショット・コピーを実行し、それが完了すると、一次サブシステム12にスナップショット・コピー操作の完了を通知する。その後、スナップショット・コピーは他のボリューム、他の遠隔サブシステム、後でアーカイブする仮想メモリ、またはスナップショット・コピーされるボリューム以外の他の任意の場所に記憶することができる。

【0021】このような機構によって、一次ボリュームへのアクセスを延期することも、スナップショット・コピーのために一次サブシステム12に記憶空間を割り振る必要もなしに、一次ボリューム32のスナップショット・コピーと同等のものが得られることを理解することが重要である。その結果、一次サブシステム12、特に一次ボリューム32の効率やアクセス可能性に悪影響を及ぼすことなく一次ボリューム32の多くのスナップショット・コピーを行うことができる。

【0022】図2を参照すると、一次サブシステム12はデータ記憶システム28と、本発明の遠隔二重コピー・プロセスとスナップショット・コピー・プロセスを管理する一次プロセッサ20とを含む。一次サブシステム12は、二次サブシステム14との間でメッセージの実際の送受信を行うための追加のデータ処理システム22を含むことが好ましい。一次データ記憶システム28はデータを記憶することができ、一次ボリューム32を含む複数のボリュームを有する典型的なマルチパス複数装置データ記憶システムとすることができる。一次データ記憶システム28は、論理アドレスを一次ボリューム32などの記憶装置の物理場所に関連づけるために任意のデータ割り振り技法を使用することができる。一次データ記憶システム28はログ構造配列（LSA）を使用することが好ましい。このような割り振り情報は一次ディレクトリ33で維持される。

【0023】二次サブシステム14は、本発明の遠隔二重コピー・プロセスとスナップショット・コピー・プロセスを管理する二次プロセッサ50と、好ましくはLSAまたは同様の配列を備えた二次データ記憶システム57を含む。二次ディレクトリ61を使用して、二次ボリ

ューム内など二次データ記憶システム57内に記憶されているデータをアドレス指定することができるようにする。また、二次サブシステム14は、書込み保留ディレクトリ121(図4に示す)を使用して、二次データ記憶システム57に記憶する更新データであるが、まだ各書込み更新セッションのための二次ディレクトリ61またはすべての書込み更新セッションのための1つの書込み更新保留ディレクトリを介してアドレス指定することができない更新データのアドレス情報をバッファリングすることができる。

【0024】一次プロセッサ20とデータ処理システム22は、一次ボリューム32上の更新データを二次サブシステム14にバックアップ・コピーする遠隔バックアップ・コピーの実行、データの記憶の監視、二次データ記憶システム57の二次ボリューム60上に記憶するために一次ボリューム32のすべての更新データを二次システム14にコピーする操作など、一次サブシステム12の内部操作を制御する。一次データ記憶システム28におけるこのようなデータ記憶監視を実現するために、プロセッサ20はPSF ESTABLISH(確立)サブシステム機能実行(PSF)コマンドなどのコマンドを、一次ボリューム32へのサブシステム・アクセス機能を有する周辺制御装置30に送ることができる。

【0025】システム相互接続40は、一次サブシステム12を遠隔二次サブシステム14に接続する通信システムである。各書込み保留セッション中に、一次サブシステム12のデータ処理システム22がデータ・システム10のための実際のデータ伝送を行う。データ処理システム22は、周知のオペレーティング・システム(OS)(図示せず)を含む。この周知のOSは、システム相互接続40を介して一次サブシステム12と二次サブシステム14との間の通信を行うために、データ・シリアルライザ及びムーバ(DSM)104及び仮想記憶通信アクセス方式(VTAM)106または同様のものを含む。システム相互接続システムに、相互接続40とVTAM24、DSM26、及びデータ・ムーバ52が含まれる。

【0026】DSM26のシリアルライザ部は、ESTABLISHED(確立)範囲によって報告されたすべてのデータ記憶システム28書込み操作に書込みシーケンス・トークンを割り当てる。「報告された」という用語は、各制御装置30に実施されたサブシステムが、更新されて記憶されたデータを遠隔で二重コピーする要求をDSM28に送ったことを意味する。このような報告には、割り当てられたタイム・スタンプとシーケンス番号を含めることができる。DSM26は、作成するメッセージにこのようなタイム・スタンプとシーケンス番号を挿入することが好ましい。DSM26は次に、これらの書込みシーケンス・トークンをVTAM24へのシステム間メッセージに入れて二次サブシステム14に送る。

DSM26は、要求、シーケンス番号、及びタイム・スタンプのテーブル(図示せず)を生成することによって要求を追跡する。

【0027】DSMのデータ・ムーバ部104は、一次サブシステム12の書き込まれた変更または更新データ・レコードを入手する。次にDSM26は、このような更新書込みレコードとそれぞれの書込みシーケンス・トークンを組み合わせて、システム相互接続40を介して二次サブシステム14内のVTAM24に送られるシステム間メッセージにする。DSM26は、一次データ記憶システム28の拡張部とみなすことができる。次にプロセッサ50が、データ・ムーバ52を使用して、受け取ったデータ・レコードを二次データ記憶システム57に記録し、書込みシーケンス・トークンをジャーナル(図示せず)にログし、二次ディレクトリ61を更新する。

【0028】前記一次サブシステム12と二次サブシステム14の両方の中のVTAM24はそれぞれ、周知のIBMシステム・ネットワーク・アーキテクチャ(SNA)通信システム及び装置のプロセッサ部とすることができる。このようなアーキテクチャについては、P. E. グリーン(Green)による論文「A Introduction to Network Architecture's and Protocols」(IBM System Journal, Vol.18, No.2, 1979)に記載されている。その他の操作及び実施態様については、米国特許第4914619号及び米国特許第4855905号も参照することができる。

【0029】VTAM24は、システム間メッセージをシステム相互接続40の周知のチャネル間(CTC)接続部またはESCON部を介して送ることができる。システム間メッセージは、一次サブシステム12から二次サブシステム14に、シャドウ(二重化)データ・エクステンション、進行中の更新を示す書込みトークン、二次データ・コピーに加える変更データのある書込みトークン、及びボリュームのスナップショット・コピー操作の実行を求めるスナップショット・コピー要求の確立を伝える。システム相互接続40は、一次サブシステムと二次サブシステムを結合する1つまたは複数のCTC接続から成る高帯域通信ネットワーク上で動作することが好ましい。T1線、T3線、またはその他の通信サービスを使用して、ESCONやその他の直接CTC接続がサポートできるよりも長い距離をサポートすることができる。

【0030】二次データ・ムーバ52が、VTAM106を介して一次サブシステム12から送られたシステム間メッセージを受け取り、書込みシーケンス・トークンを抽出して維持し、受け取った一次サブシステム12更新データを二次データ記憶システム19に記録する。データ・ムーバ52は、二次データ記憶システム57の拡張部と見なすことができる。一次サブシステム12と二

次サブシステム14を同じ物理場所に配置して、VTAMデータ通信リンクを不要にすることもできることを理解されたい。すなわち、一次サブシステム12と二次サブシステム14との間の物理的間隔と、2つのサブシステム間のデータリンクのタイプは、本発明の実施には限定されない。

【0031】図3に、一次ボリューム32と二次ボリューム60上にそれぞれデータを記憶する操作と、一次ディレクトリ33と二次ディレクトリ61をそれぞれ更新する操作とを制御する制御装置30及び58の一実施形態を示す。各制御装置30及び58は、キャッシュ84、NVキャッシュ86と、マイクロプロセッサ88と、ランダム・アクセス・メモリ(RAM)90と、NVRAM92と、電源接続94とを含む。制御装置30及び58の構成要素は当技術分野で周知であり、詳細には説明しない。一次ディレクトリ33と二次ディレクトリ61は、容易にアクセスすることができるようそれぞれの制御装置30及び58のNVRAM92の不揮発性メモリに記憶することができることに留意されたい。

【0032】図4に、参照番号120で示された、ログ構造配列(LSA)形式に構成された一次ディレクトリ33と二次ディレクトリ61の好ましい実施形態を示す。本明細書では「データ単位」という用語を使用して、所与のファイル内の所与のレコードの所与のバイトを示す。本発明では、一次ボリューム32の更新データのすべてのコピーが、二次サブシステム14の二次ボリューム60にコピーされて、PPRCの対が二重化状態に維持される。

【0033】具体的には、制御装置30及び58は、それぞれのデータ記憶システム28及び57に論理レコードを書き込む。それぞれのデータ記憶システム28及び57で受け取った更新レコードは、揮発性キャッシュ112またはNVキャッシュ113(図2)に一時的に記憶して、圧縮(必要な場合)とそれぞれの物理ディスク装置(DASD)への固定記憶が行われるのを待つことができる。一次サブシステム12には、データは論理ユニットの論理トラック単位で記憶することが好ましい。二次サブシステム14は更新データを論理トラック・ボリュームに記憶することもできるが、このような二重コピー・データへのアクセスは主として回復や置換のために行われるため、この要件は二次サブシステムにとって重要ではないと考えられる。

【0034】LSA120で表された一次ディレクトリ33及び二次ディレクトリ61は、それぞれの一次ボリューム32及び二次ボリューム60についてアドレス可能論理トラックに可変長の論理レコードを記憶することができる。このような論理トラックは、単一のブロックまたは単位として記憶することも、複数のトラック部分として、すなわち断片化して記憶することもできる。最小の断片は1データ・バイトである。

【0035】LSA120は、N+1の物理ディスク装置(DASD)のための、言い換えると、データ記憶システムのボリュームのための情報を維持する。一次ボリューム32及び二次ボリューム60を表す各ボリューム100は論理的に、LSAセグメント・コラム104と呼ぶ複数の連続してアドレス指定可能な領域に分割されている。各LSAセグメント・コラム104は1シリンダのトラックで構成することができる。それぞれのボリューム32及び60内の同様の物理的地址を有する各LSAセグメント・コラム104の総体が、両向き矢印102で示されている1つのLSAセグメントを構成する。各LSA内のLSAセグメント・コラム104のうちの1つに、LSAセグメント内の他のすべてのLSAセグメント・コラムの誤り検出冗長要素を含めることができる。

【0036】構成された位置実施形態では、誤り検出冗長要素は奇偶数などのパリティである。様々なLSAセグメント102内のパリティLSAセグメント・コラムは、物理ボリューム32及び60のうちの異なるボリュームにあることが好ましい。カウント・キー・データ(CKD)または固定ブロック(FB)形式に形式設定されたデータを記憶する。各論理トラックは、1組の圧縮ファイルとしてLSAセグメント・コラム104のうちの1つの中に完全に記憶することができる。各LSAセグメント・コラム104に多くの論理トラックを記憶することができる。各論理トラックまたはその各記憶断片の物理記憶場所は更新のたびに変わる。

【0037】LSAディレクトリ120は、論理トラック・フィールド122と、物理アドレス・フィールド124と、長さフィールド126とを含み、各論理トラックごとに1つの項目を維持する。フィールド122には、別々に記憶されている各論理トラック断片すなわちバイト、バイト・クラスタ、セクタなどの識別情報を含む論理トラック識別情報(論理アドレス)が入る。フィールド124には、別々の各トラック断片の識別情報を含む論理トラックが記憶されている物理アドレスが入る。フィールド126は、ボリューム100上の物理セクタ(図示せず)に記憶されている論理トラックと各論理トラック断片の長さに関する。論理トラックにアクセスするには、まずLSAディレクトリ120にアクセスし、次にそれぞれのLSAセグメント・コラム内の物理セクタにアクセスする必要がある。LSAディレクトリ120全体を不揮発性ランダムアクセス制御メモリNVRAM116に記憶してアクセスを容易にすることが好ましい。

【0038】LSA120へのデータの書込みは、以下に説明する手続きに従って行われることが好ましい。LSA120への各書込みは、LSAセグメント・コラム104のうちのアドレス指定された1つの列への論理トラック106または断片の書込みである。このようなア

ドレス指定されたLSAセグメント・コラムは、そのようなアドレス指定された論理トラックまたは断片のデータを現在記憶中のLSAセグメント・コラムではない。論理ボリュームの論理トラックまたは論理トラックのいずれかの断片にデータがない場合、そのトラックまたは断片にはそれぞれLSA120内のデータ記憶空間が割り振られない。LSAに記憶されているデータが更新された場合、その論理トラックの第2のコピーが記録され、元のデータのアドレス可能性が取り除かれる。たとえば、領域108及び110が更新データを表すとする。この更新データを記録するには、それぞれの論理トラック108またはその断片の置換データを新しいデータ記憶場所に書き込む。交互的に空き空間コレクタと呼ばれている通常のデータ記憶スペース回復技法によって、LSAセグメント102内の記憶空間を再生する。LSAセグメントにデータを記録している間は、前述のパリティLSAセグメントは記録されない。N個のLSAセグメント・コラムがすべていっぱいになると、パリティLSAセグメント・コラムN+1にパリティまたはその他の誤り検出冗長要素が入れられる。当然ながら、所与のLSAセグメント内に記憶されているデータのために希望する場合には、パリティを先に書き込むこともできる。更新された論理トラック断片を記憶するNVキャッシュ113によって、DASDに対するこのような更新データの書き込みを据え置くことができる。

【0039】図4を参照すると、書き込みディレクトリ121はデータ識別情報フィールド128と、LSAアドレス・フィールド130と、シーケンス及びタイム・スタンプ・フィールド132とを含む。書き込みディレクトリ121は、二次データ記憶システム57に記憶する更新データであるが、まだ各書き込み更新保留セッションのための二次ディレクトリ61またはすべての書き込み更新セッションのための1つ書き込み更新保留ディレクトリを使用してアドレス指定することができない更新データのアドレス情報をバッファリングする。フィールド128によって更新する特定のデータが識別される。フィールド130は、更新データを記憶する特定のLSAアドレスに関係する。フィールド132は、ボリューム100、すなわち二次ボリューム60内でデータを更新する順序を判断するためのシーケンスとタイム・スタンプに関係する。

【0040】本発明の動作例について、図5を参照しながら以下に説明する。最初に、ホスト装置（図示せず）が一次サブシステム12の一次ボリューム32のスナップショット・コピーを要求する（ステップ300）。一次サブシステム12は一次ボリューム32のスナップショット・コピーを求める要求を受け取る（ステップ302）。一次サブシステム12は、遠隔コピー・セッションが進行中であるかどうかを判断する（ステップ304）。遠隔コピー・セッションが進行中の場合、一次サ

ブシステム12は、現時点ではスナップショット・コピー操作を行うことができないことを示すメッセージをホスト装置に送る（ステップ306及び308）。

【0041】遠隔コピー・セッションが進行中でない場合、一次サブシステム12は、二次ボリューム60のスナップショット・コピーを行う二次サブシステム14にスナップショット要求を中継または転送する（ステップ310）。次に、二次ボリューム60のスナップショット・コピーが、後でアーカイブされる仮想メモリまたは二次ボリューム60以外の他の記憶場所のローカルまたは遠隔場所の第3のボリュームに記憶される（ステップ312）。スナップショット・コピーのために必要な記憶空間を最小限にするために、現行スナップショット・コピーが前のスナップショット・コピーを上書きすることが好ましい。

【0042】スナップショット・コピー操作が完了すると、二次サブシステム14は二次ボリュームのスナップショット・コピーの完了を検証するために一次サブシステム12に検証信号を送る（ステップ314）。検証信号には、スナップショット・コピーの場所を示す追加の情報も含めることができる。その後、一次サブシステム12は検証信号をホスト装置に中継または転送し（ステップ316）、このプロセスを完了する（ステップ318）。

【0043】まとめとして、本発明の構成に関して以下の事項を開示する。

【0044】（1）ボリュームのスナップショット・コピーを行う装置であって、一次記憶サブシステムと二次記憶サブシステムとを含み、前記一次記憶サブシステムは、（a）少なくとも1つの一次ボリュームを有する一次データ記憶装置と、（b）前記一次ボリュームの少なくとも一部のスナップショット・コピーを行う要求を二次記憶サブシステムに中継する一次処理手段とを含み、前記二次記憶サブシステムは、（c）前記一次ボリュームの複製コピーである二次ボリュームを有する二次データ記憶装置と、（d）前記要求にตอบสนองして、前記二次ボリュームに対応する部分のスナップショット・コピーを行うことによって前記一次ボリュームの前記少なくとも一部のスナップショット・コピーと同等のものを実行する二次処理手段とを含む装置。

（2）前記二次ボリュームが前記一次ボリュームの複製コピーになるように、前記一次ボリュームの前記二次ボリュームへのバックアップ・コピー操作を実行する手段をさらに含む、上記（1）に記載の装置。

（3）前記バックアップ・コピー操作が進行中でない場合にのみ、前記一次処理手段が前記要求を中継する、上記（2）に記載の装置。

（4）前記スナップショット・コピーが第3の遠隔ボリュームに記憶される、上記（1）に記載の装置。

（5）前記二次ボリュームの前記少なくとも一部の前記

スナップショット・コピーが前記二次記憶サブシステムの仮想メモリに記憶される、上記(1)に記載の装置。

(6) 前記一次記憶サブシステムが、前記一次ボリューム上のデータの物理場所を仮想アドレスに関連づける一次ディレクトリ情報を含む、上記(1)に記載の装置。

(7) 前記一次ディレクトリ情報がログ構造配列形式である、上記(6)に記載の装置。

(8) 前記二次記憶サブシステムが、前記二次ボリューム上のデータの物理場所を仮想アドレスに関連づける二次ディレクトリ情報を含む、上記(1)に記載の装置。

(9) 前記二次ディレクトリ情報がログ構造配列形式に構成されている、上記(8)に記載の装置。

(10) 前記スナップショット・コピーが完了すると、前記二次記憶サブシステムが前記スナップショット・コピー操作の完了を示す信号を前記一次記憶サブシステムに送る、上記(1)に記載の装置。

(11) ボリュームのスナップショット・コピーを行う方法であって、一次記憶サブシステムの一次ボリュームの少なくとも一部のスナップショット・コピーを行う要求を前記一次記憶サブシステムで受け取るステップと、前記要求を前記一次記憶サブシステムから、前記一次ボリュームの複製である二次ボリュームを有する二次記憶サブシステムに中継するステップと、前記二次ボリュームの対応する部分についてスナップショット・コピーを行うことによって前記一次ボリュームの前記少なくとも一部のスナップショット・コピーと同等のものを生成するステップとを含む方法。

(12) 前記二次ボリュームが前記一次ボリュームの複製になるように、前記一次記憶サブシステムが、前記二次ボリュームへの前記一次ボリュームのバックアップ・

コピー操作を行う手段を含む、上記(11)に記載の方法。

(13) 前記バックアップ操作が進行中でない場合に前記生成ステップが行われる、上記(11)に記載の方法。

【図面の簡単な説明】

【図1】本発明によるデータ記憶システムの全体構造図である。

【図2】本発明によるデータ記憶システムの構成要素を示すブロック図である。

【図3】本発明によるデータ記憶システムの制御装置を示す図である。

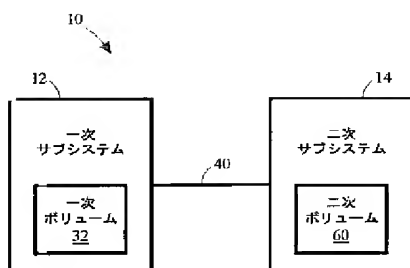
【図4】本発明によるログ構造配列(LSA)を採用したデータ記憶システムを示す図である。

【図5】本発明による二次ボリュームのスナップショット・コピー操作を示す流れ図である。

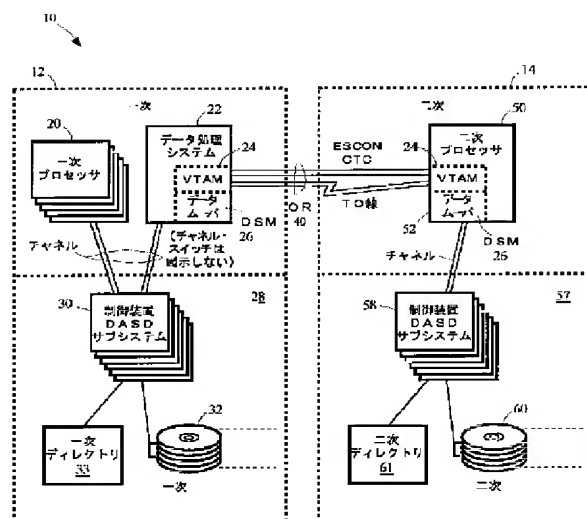
【符号の説明】

- 10 データ・システム
- 12 一次サブシステム
- 14 遠隔二次サブシステム
- 20 一次プロセッサ
- 22 データ処理システム
- 28 データ記憶システム
- 32 一次ボリューム
- 33 一次ディレクトリ
- 40 通信インタフェース
- 50 二次プロセッサ
- 57 二次データ記憶システム
- 60 二次ボリューム
- 61 二次ディレクトリ

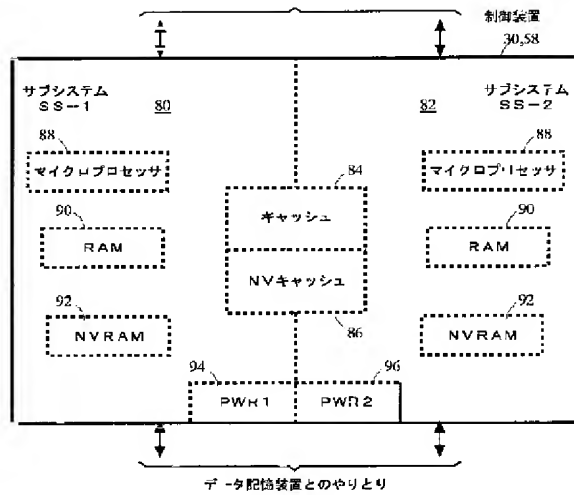
【図1】



【図2】

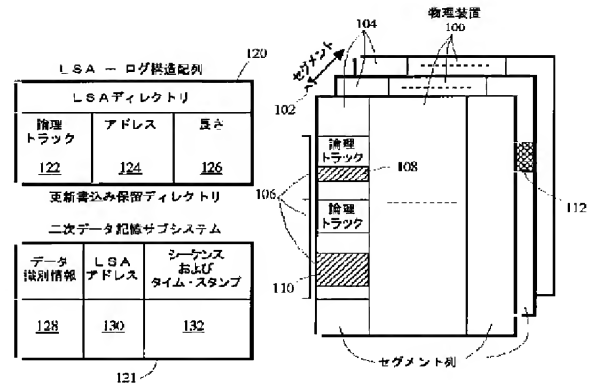


【図3】

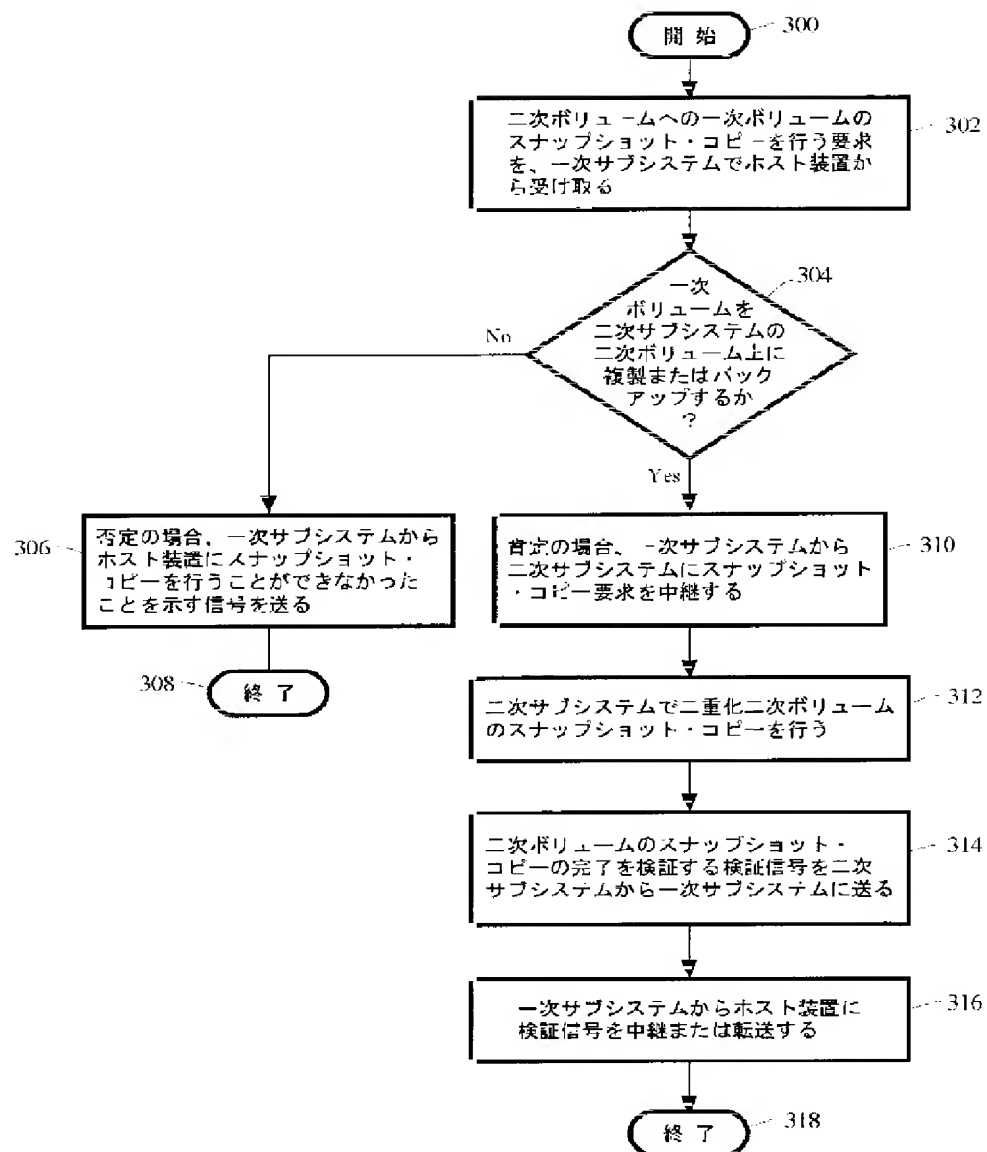


NVキャッシュ 不揮発性キャッシュ
 NVRAM 不揮発性RAM
 RAM ランダム・アクセス・メモリ
 PWR 電源
 SS サブシステム

【図4】



【図5】



フロントページの続き

(72)発明者 デービッド・アラン・サールズ
アメリカ合衆国80127 コロラド州リトル
トン サウス・ネルソン・ストリート
4817